# IN THE CLAIMS:

Claims 1 and 10 are amended herein. All pending claims and their present status are produced below.

1.    (Currently Amended) A <u>computer implemented</u> data quality system <u>executed by a computer server</u> for matching input data across record pairs of data <u>stored in and retrieved from the computer server</u>, the system comprising:

training means, comprising, for a training data set comprising a portion of the data set:

means for matching pairs of records of the training data set by matching pairs of data fields between each pair of records to generate for each pair of records a similarity vector comprising a plurality of similarity scores, each similarity score associated with a pair of data fields;

means for applying a set of rules to the data fields of the records and for applying weights to the plurality of similarity scores of each similarity vector for the training data set to calculate an overall similarity score for each pair of records in the training data set; and

means for determining, in response to user feedback with respect to the training data set, a set of adjusted rules and weights;

for a non-training data set comprising unmatched pairs of records of the data set, means for matching the pairs of records of the non-training data set by matching pairs of data fields between each pair of records to generate similarity vectors for the pairs of records of the non-training data set;

means for applying the set of adjusted rules and weights to similarity scores of the similarity vectors for the pairs of records of the non-training data set to calculate an overall similarity score for each pair of records in the non-training data set.

2.    (Canceled)

3.      (Previously Presented) A system as claimed in claim 1, wherein the means for matching pairs of records of the training data set comprises means for executing string matching routines on pre-selected field pairs of the records.

4.      (Previously Presented) A system as claimed in claim 3, wherein a selected string matching routine comprises means for determining an edit distance indicating the number of edits required to change from one value to  another value.

5.      (Previously Presented) A system as claimed in claim 3, wherein a selected string matching routine comprises means for comparing numerical values by applying numerical weights to digit positions.

6.      (Previously Presented) A system as claimed in  claim 1, wherein the means for matching pairs of records of the training data set comprises means for generating a vector value between 0 and 1 for each  pair of data fields in a  pair of records.

7.      (Previously Presented) A system as claimed in  claim 1, wherein the  means for matching pairs of records of the training data set comprises record scoring means for converting  a similarity vector into  an overall similarity score representing overall similarity of the fields in each record pair.

8.      (Canceled)

9.      (Previously Presented) A system as claimed in claim 7, wherein the record scoring means comprises means for computing scores using an artificial intelligence technique to deduce from examples given by the user an optimum routine for computing the overall similarity score from the vector.

10.     (Currently Amended) A system as claimed in claim 9, wherein the artificial intelligence technique used is ~~eased~~ case based reasoning (CBR).

11.     (Original) A system as claimed in claim 9, where the artificial intelligence technique used comprises neural network processing.

12.     (Previously Presented) A system as claimed in claim 1, further comprising a pre-processing means for reformatting the data set, further comprising a standardization module comprising means for transforming each data field into one or more target data fields each of which is a variation of the original.

13. (Previously Presented) A system as claimed in claim 12, wherein the standardization module comprises means for splitting a data field into multiple field elements, converting the field elements to a different format, removing noise characters, and replacing elements with equivalent elements selected from an equivalent table.

14. (Previously Presented) A system as claimed in claim 12, wherein the pre-processing means comprises a grouping module comprising means for grouping records according to features to ensure that all actual matches of a record are within a group, and wherein the means for matching pairs of records of the training data set comprises means for comparing records within groups only.

15. (Previously Presented) A system as claimed in claim 14, wherein the grouping module comprises means for applying labels to a record in which a label is determined for a plurality of fields in the record and records are grouped according to similarity of the labels.

16. (Original) A system as claimed in claim 15, in which a label is a key letter for a field.

17. (Previously Presented) A system as claimed in claim 12, wherein the system further comprises a configuration manager comprising means for applying configurable settings for the preprocessing means and for the means for matching pairs of records of the training data set.

18. (Previously Presented) A system as claimed in claim 7, wherein the system further comprises a tuning manager comprising means for refining, according to user inputs, operation of the record scoring means.

19. (Previously Presented) A system as claimed in claim 18, wherein the tuning manager comprises means for using a rule-based approach for a first training run and the artificial intelligence approach for subsequent runs.